DOCUMENT RESUME

ED 397 118 TM 025 252

AUTHOR Finzer, Bill; Chen, Eugene
TITLE Components for Data Analysis and Statistics. The DataSpace (TM) Project.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Apr 96
NOTE 12p.; Paper (poster) presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 9-12, 1996). Illustrations may not reproduce well.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Computer Assisted Instruction; Computer Networks; Computer Software; Curriculum Development; *Data Analysis; Graphs; High Schools; Mathematics Curriculum; *Mathematics Instruction; *Statistics; Tables (Data)

ABSTRACT

The goal of the DataSpace project is to develop a computer learning environment for data analysis and statistics and associated materials for high school mathematics. DataSpace software will provide a concrete metaphor for data, along with interactive visual explanations of statistical concepts. Users will be able to change data smoothly and continuously and construct and study their own statistical measures and graphs. When the suite of software is completed, it will draw students into learning and using data analysis and statistics. Illustrations of software screens depict the way in which DataSpace will provide dynamic tables and graphs. The National Council of Teachers of Mathematics has recommended increased classroom emphasis on data analysis and statistics in the rest of the mathematics curriculum. The DataSpace project will eventually provide three units of material sufficient for a one-semester statistics course or as part of an integrated mathematics curriculum. The software will be an important tool for the statistics teacher and will take advantage of the networking structures being put into schools. (Contains seven illustrated screens.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 118

Components for Data Analysis and Statistics

The DataSpace™ Project[1]

A poster for "Scriptable Components: A New Approach to High Impact Learning Technologies"

to be presented at AERA '96, New York, by

Bill Finzer, Educational Technology Director, and

Eugene Chen, Software Designer

Key Curriculum Press

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
WILLIAM F. FINZER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Educational Objective of the DataSpace Software

The goal of the DataSpace project is to develop a computer learning environment for data analysis and statistics and associated curriculum materials for high school mathematics. DataSpace software will

- provide a concrete metaphor for data—"cases" and "collections"—
- provide interactive visual explanations of statistical concepts
- allow the user to change data smoothly and continuously to observe the effects of those changes on all dependent graphs and analyses
- encourage users to construct and study their own statistical measures and graphs
- integrate resampling techniques with statistical analysis
- integrate simulation with data analysis
- deal with the timely, real world data available in large and remote datasets.

Pedagogical and Technological Obstacles

Commercially available statistics packages aim at statistically literate practitioners rather than learners. Graphing calculators can do a surprising amount of statistics, but are still clumsy at helping students understand and appreciate what is going on. While good progress has been made in rethinking the teaching of data analysis and statistics, much less progress has been made in providing useful learning tools.

[1]This material is based on work supported by the National Science Foundation under award number III-9400091. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

TM025251

BEST COPY AVAILABLE

How can we provide software and curriculum that involve the learner in constructing their own understanding of data analytic and statistical concepts?

How can these materials best be made engaging, illuminating, useful for real tasks, and fun?

How can we complete the design and implementation of an ambitious piece of educational software with limited people and time resources?

Design solution

We are designing (and beginning to implement using OpenDoc™ and OpenDoc Framework[1]) a suite of software components and curriculum units that work together to draw students into learning and using data analysis and statistics. Components make sense both from an end-user point of view in which students can mix and match tool components from different environments, and from a development point of view in which they provide factorization and reusability.

The Cases and Collections Metaphor

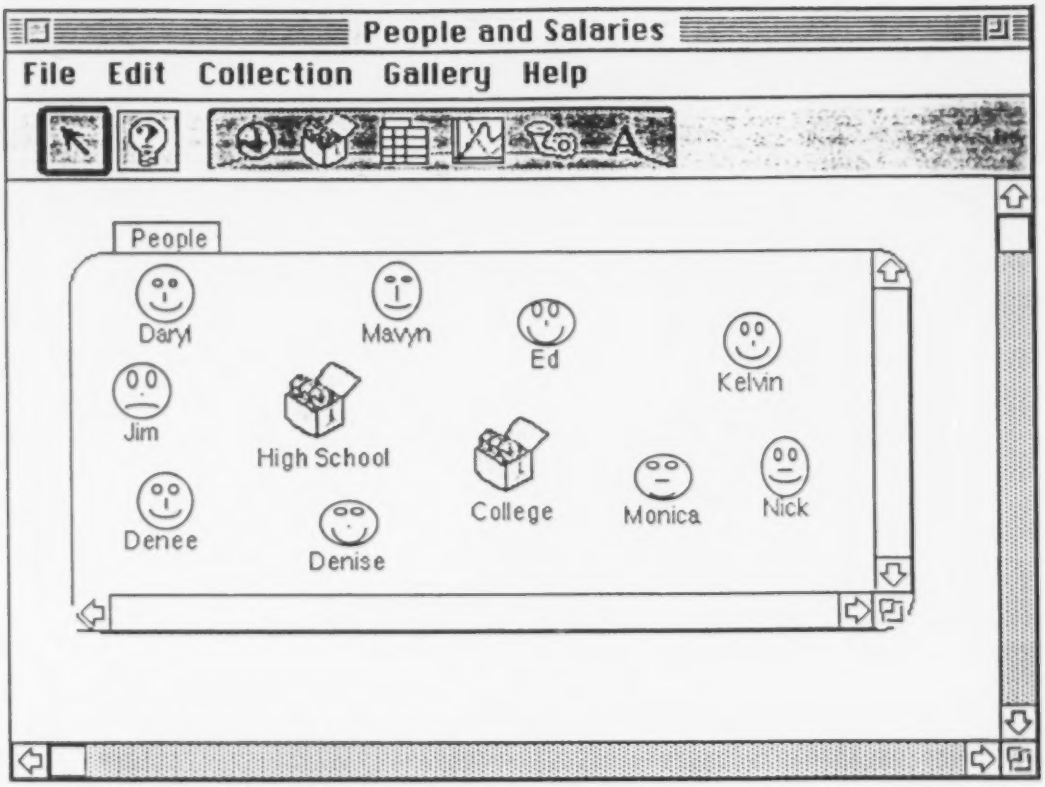
Understanding how to represent data poses a learning hurdle for students as they strive to make sense of many mathematical and physical concepts[2]. Typically educational applications represent data in tabular form with rows holding individual cases and columns holding attributes. DataSpace, taking its lead from Chris Hancock's TableTop, encourages a case-centric representation of data in which individual cases appear as icons in a collection where they may be moved around, grouped into subcollections, made to position themselves according to a formula, or given an appearance that reflects the data contained within.

The figure below shows a DataSpace document containing a collection of people with subcollections in which the user has placed people who finished high school and others who went to college, but not beyond. There are some left over people who don't belong in either category. The appearances of the people reflect other attributes: face shape for urban, suburban, or rural; eye shape for kind of work, and income level for breadth of smile.

[1] Currently we are using OpenDoc for the Mac™ OS, ©1995 Apple Computer. OpenDoc Framework is a multi-platform development framework that should allow us to ship the software under both Macintosh and Windows™.

[2] Hancock, Chris, "The Medium and the Curriculum: Reflections on Transparent Tools and Tacit Mathematics," in *Computers and Exploratory Learning*, eds: diSessa, A., Hoyles, C., Now, R., and Edwards, L., Heidleberg/Springer-Verlag, 1995.

BEST COPY AVAILABLE

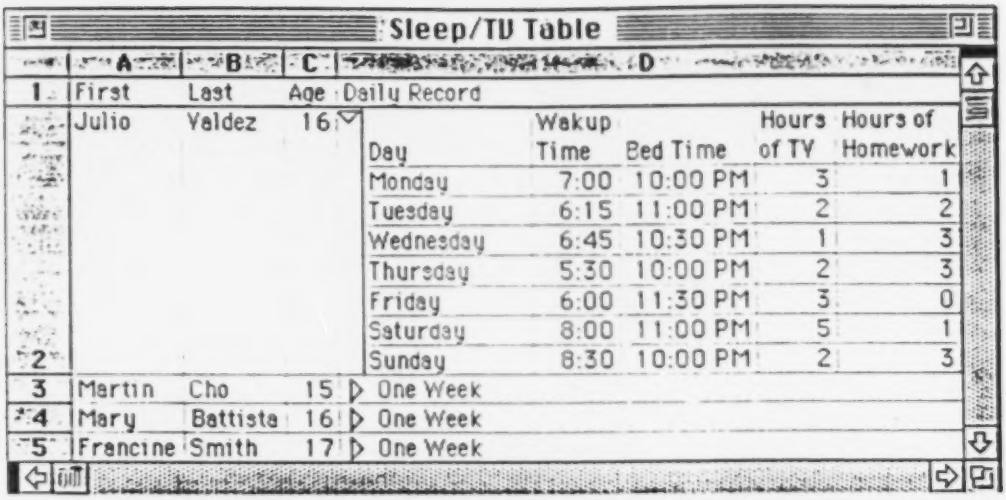


If, as seems likely, a concrete metaphor for cases and collections proves helpful to students as they work with data, it would beneficial to bring the DataSpace collection component into other computer learning environments which deal with data. This component will define the interface for assigning representative appearances to cases of data. In addition, the positional coordinates of the icons can be linked to the underlying data to allow direct manipulation of the data by dragging the icon. The caption under the icon can likewise provide an edit box for one field of the data or might present some transformation on the data, for instance, formatting an English sentence from fields of the data.

Components facilitate sharing metaphors among different learning environments, increasing opportunities for transfer between and synergy among concept domains.

Tables Within Tables

Near to the heart of any data analysis and statistics application are table and graph components. In DataSpace tables provide a form of data entry, supplementing data entry through individual cases, and they display the output of statistical tests and measures. Because DataSpace permits values that are other collections, DataSpace table components will be able to display tables within tables as shown here in this hypothetical example of students collecting a week's worth of sleep, TV, and homework data.



Of course it will cost the DataSpace project significant design, programming, and testing resources to create these hierarchical tables. We have much greater incentive to build them if we believe they can be reused in other settings. And other projects will be more likely to consider using powerful data structures if display and editing parts for these data structures are available to them.

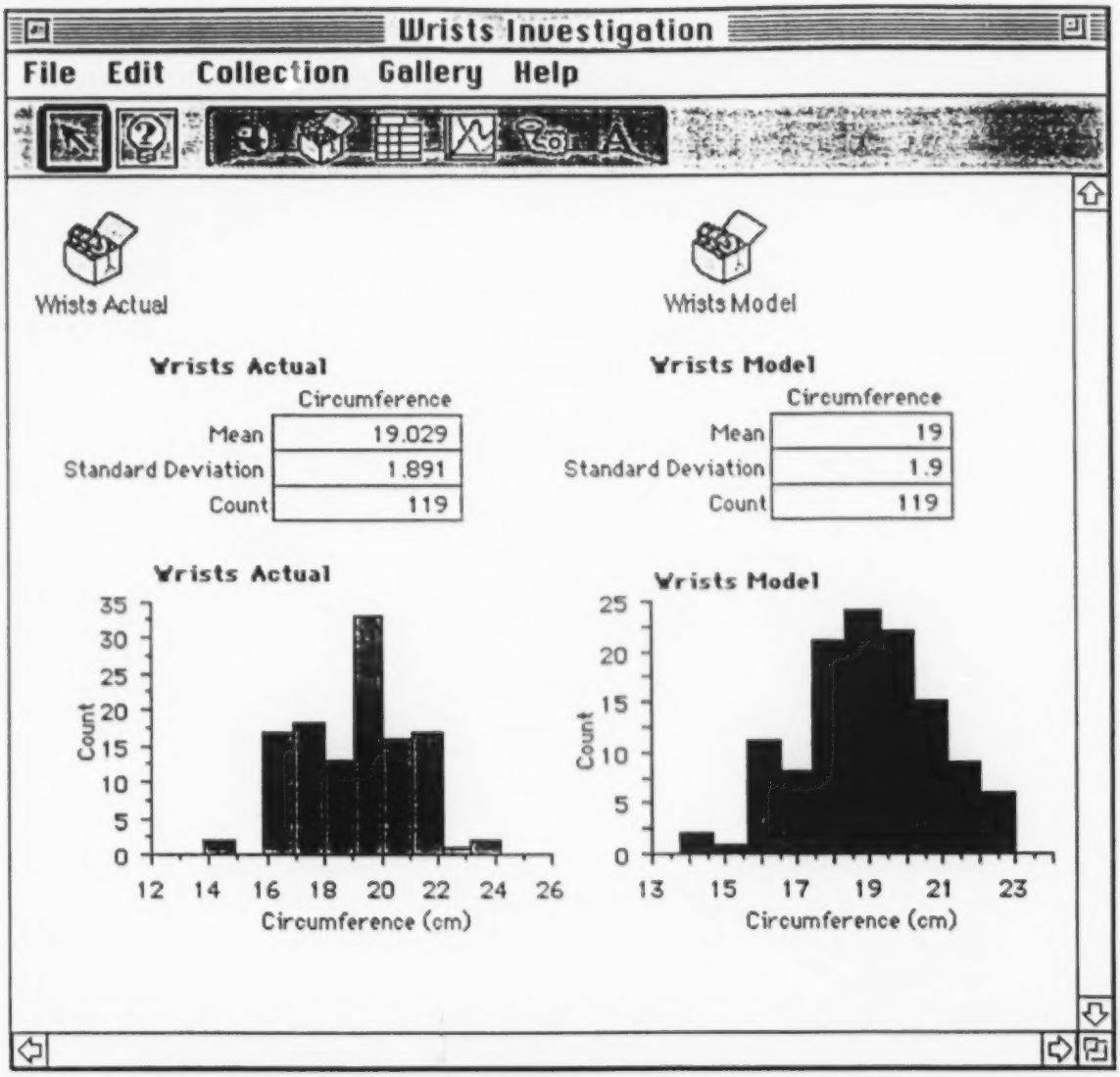
As a community of researchers and developers attempting to understand and change educational practice, we are severely limited by the difficulty of building on each other's work. Component technology gives us a much-needed mechanism for leveraging new work on old and for collaboration between projects.

Virtual Collections

Although electronic dissemination media such as the Internet and CD-ROMs have made case data increasingly available, much, if not most, data accessible in classrooms is summary data in which the individual cases have been lost. DataSpace tables allow users to enter data into summary tables for the purposes of graphing and performing appropriate statistical tests. In addition, by specifying the type of distribution, DataSpace users can frequently generate a "virtual" collection of case data from the summary data. This ability to move from summary to case data makes it possible to look at the many possible kinds of distributions from which the summary data might have sprung.

With virtual collections users can easily compare actual with simulated data. Shown here are two collections of data, one gathered from measurements and the other a normal distribution with nearly the same mean and standard deviation as the actual data.[1]

[1] In actuality, both of these collections are simulated. The one on the left is a Poisson distribution while the one on the right is normal.



Though we are bombarded by random phenomena and variation—ambient noise, fluttering tree leaves, motion of crowds and traffic—as learners we get far too few opportunities to study variation in a structured setting. Virtual collections and other features of DataSpace should help students gain experience with and intuition about these phenomena.

DataSpace components such as virtual collections should prove useful in other settings in which starting from summary data is fundamental.

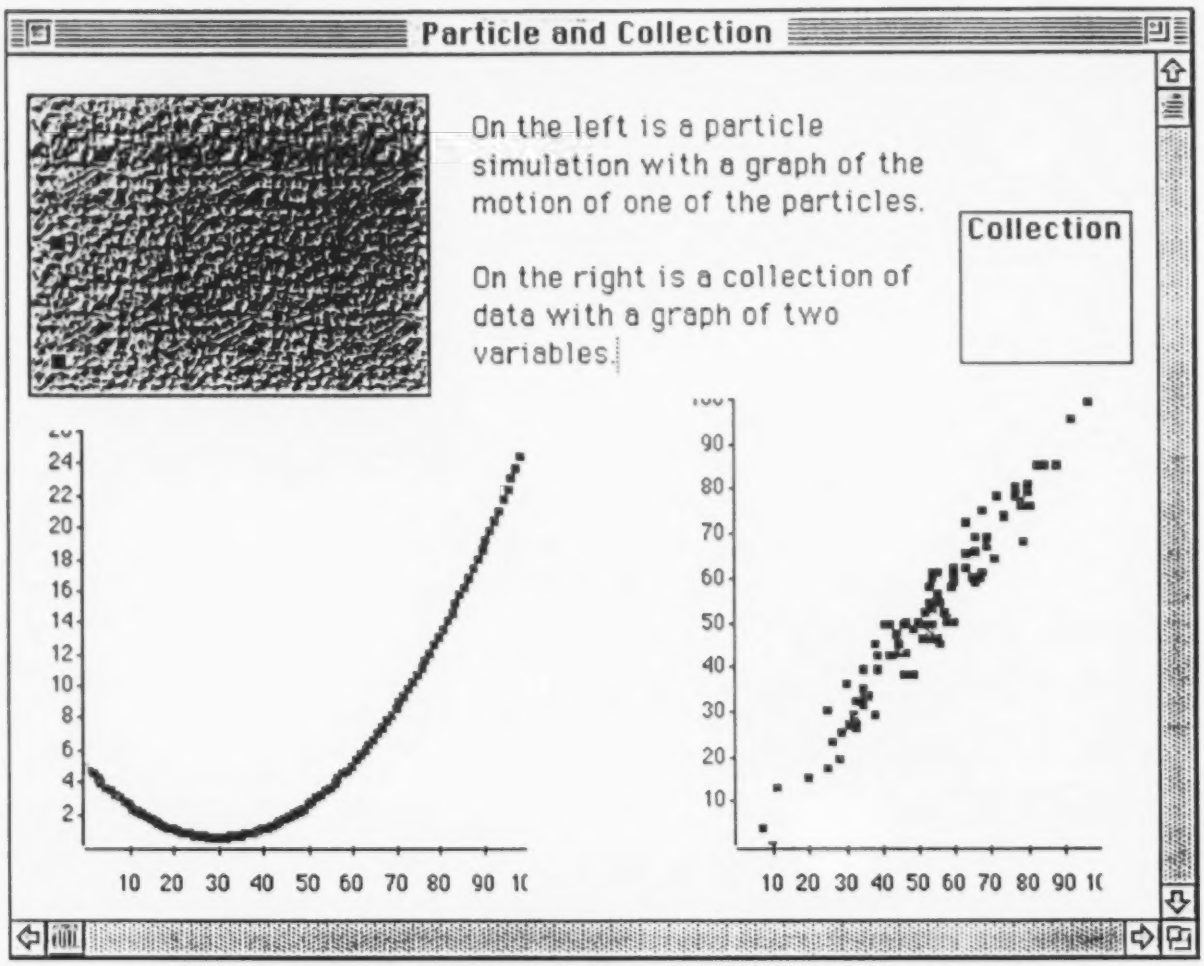
The Importance of Graphs

Graphs are both the paradigmatic illustration of the benefit that could accrue with reusable component technology and of the difficulty of providing such benefit. Nearly every mathematics and science educational software development project has to start over implementing graphs. Reusable graphs could free up resources.

BEST COPY AVAILABLE

But at the same time, every project has different requirements for graphs, especially different forms of user-interaction with graphs.

In the DataSpace project we have begun implementing OpenDoc parts, including a prototype scatterplot part that can plot data both from prototypes of DataSpace collections and from a particle motion simulation developed by Jeremy Roschelle of the SimCalc project. The illustration here shows prototypes from both projects living as parts in the same document.

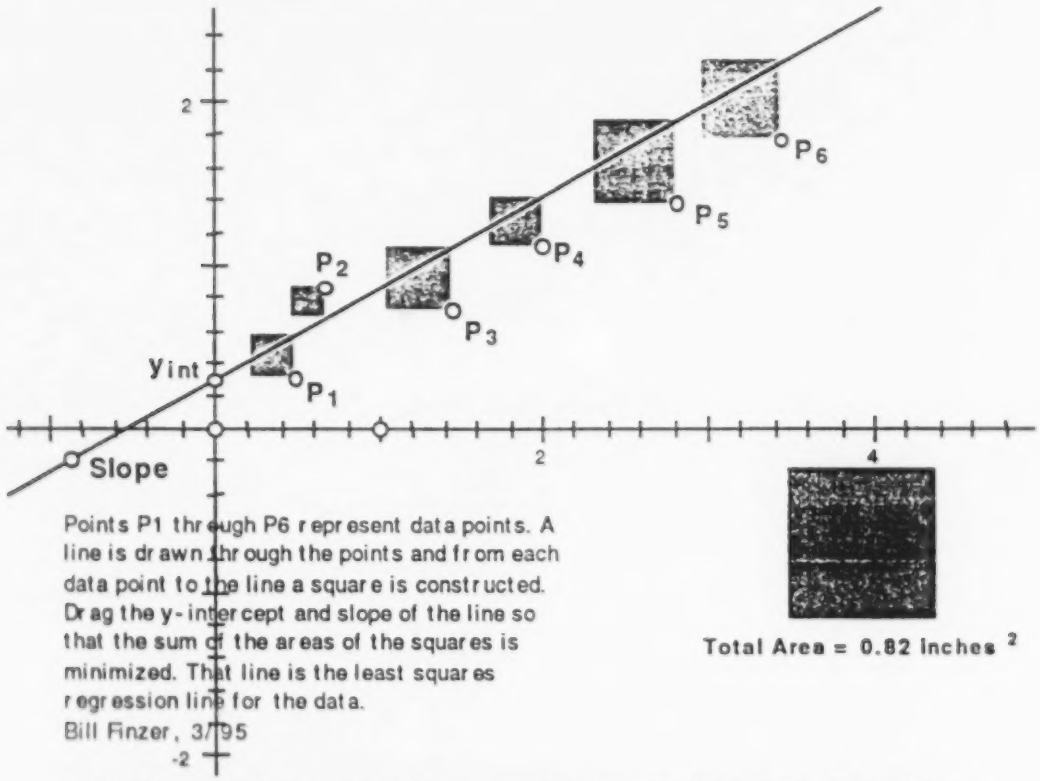


DataSpace graphs will be dynamic in the sense that users can change data by clicking and dragging on graph elements, and the effects of these changes are manifest dynamically during the drag. We suspect that these dynamic capabilities will be useful in other learning environments besides DataSpace.

Reusing graphs among projects will require more than standards. It will require a commitment to collaboration so that user interaction with graphs can be tuned to the needs of each of the individual projects. A standard graph interaction suite may eventually emerge, but it will take years of working together to get there.

A Dynamic Geometry™ Component to Visualize Statistical Concepts

In writing the original DataSpace proposal, we needed an illustration of a least squares regression line. Because we were already involved in research relating to The Geometer's Sketchpad[1], we created the illustration using that software. The result, shown here, is a sketch in which you can drag data points or drag a line and, in either case, observe the effect of the drag on the total squares of deviations. Once having played with this sketch, you can never again be mystified about the term "least squares."


2
P6
P5
P4
P2
P3
yint
P1
Slope
2
4
Points P1 through P6 represent data points. A line is drawn through the points and from each data point to the line a square is constructed. Drag the y-intercept and slope of the line so that the sum of the areas of the squares is minimized. That line is the least squares regression line for the data.
Bill Finzer, 3/95
-2
Total Area = 0.82 Inches 2

The Geometer's Sketchpad has proven itself useful in modeling a surprising variety of concepts from mathematics and science. As part of our work in the DataSpace project, we have created visualizations of the mean, median, standard deviation, the chi square test, the correlation coefficient, and normal distributions. An open component for modeling geometry can be used by other projects to tailor non-standard graphs—either programmatically or by their end-users.

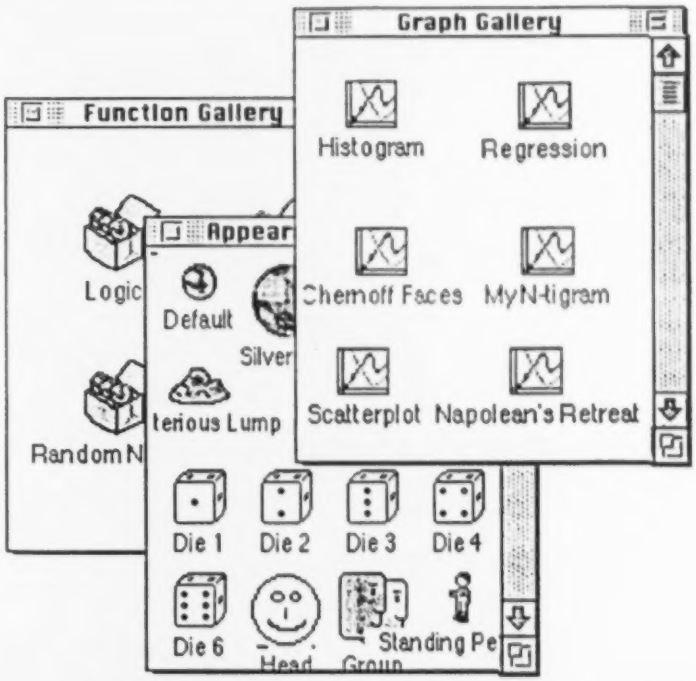
DataSpace's requirements for a dynamic geometry component are stiff—the component should accept data from other components and allow the user to change that data during a drag. The payoffs are high—the insight gained by watching a system change continuously during a drag seems qualitatively different (and greater) than that gained by comparing the system before and after a parameter change.

The world of math and science education software needs a reusable dynamic geometry modeling component.

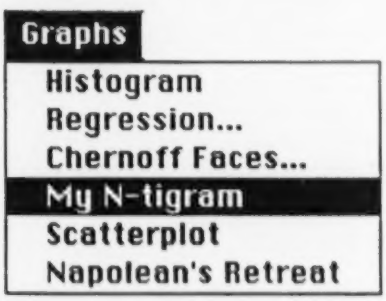
[1] Supported by the National Science Foundation under award number III-9203421.

Galleries of Components

Collections in DataSpace can act as a repository for more than just cases of data. They are also used to hold onto frequently used icons, formulas, graphs, and tables. When used in this way, we call the collection a *Gallery* and present it to the user as a palette, i.e. the collection appears in a window which is always on top. This provides for more direct manipulation of program entities than do menus or dialogs.



The Appearance Gallery holds icons which can be used to customize the way things look. The Function Gallery provides browsing of numerous functions, avoiding the need to remember and type syntax. In DataSpace, graphs are thoroughly configurable. The Graph Gallery will contain graphs which have been preconfigured to provide the canonical set (histogram, scatterplots etc.). These can be dragged into the user's document, where they may be used as is or modified to fit the user's needs.



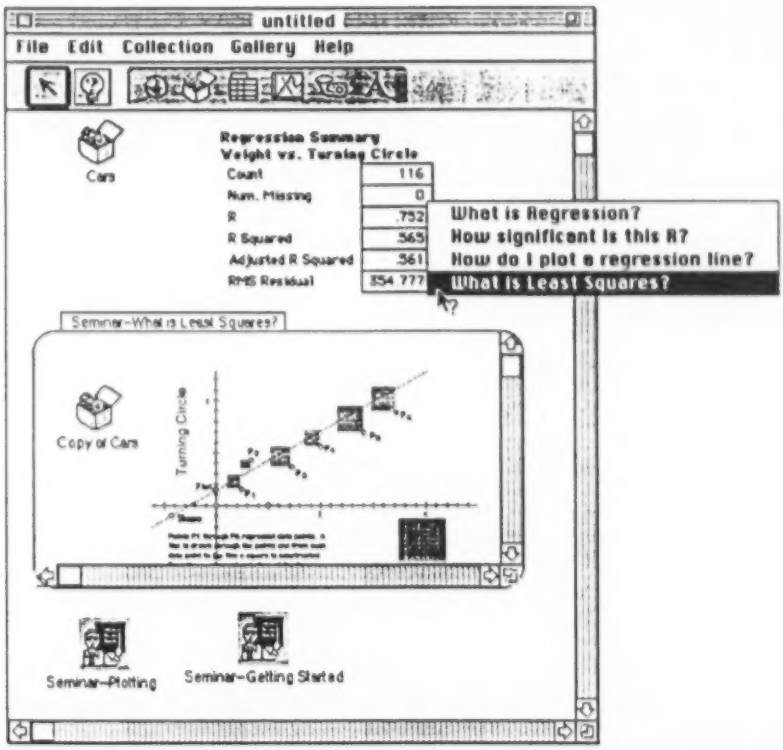
Galleries allow the user to extend the program. For instance, in DataSpace, the available graphs are all kept in the Graph Gallery and appear as choices from a pull-down men. When the user has created a graph of general utility, she can add it to the graph gallery. Thereafter, these graphs appear in the program alongside the "official" ones—constructed graphs are just as real as the official graphs.

DataSpace galleries provide a general solution to user cataloging and retrieval of frequently used components. By being sensitive to the presence of galleries, programs open themselves to user extension.

Galleries will be implemented as fairly generic containers for any component, allowing nesting via DataSpace's collection metaphor. They will define a standard mechanism for displaying, saving, cloning, and drag and drop access to these components.

Galleries Can Help

We plan to have an extensive help system for DataSpace which will provide explanations not only on the operation of the program but also serve as a context sensitive reference manual for curricular activity. An example of this is the Least Squares sketch mentioned previously. We store these explanations in collections which we call seminars. Naturally there is a Seminar Gallery, where users can browse these explanations. Seminars not only give access to the visual explanations, they allow the explanations themselves to become first class objects in the document.



Students can point at any DataSpace object using the Question Tool. A menu will pop-up displaying questions which might pertain to that object. The answer to the question the student chooses is a seminar which is copied into his document. This seminar goes beyond static help text by using a copy of student's own data within the example and allowing him to interact with the explanation. Because Seminars are ordinary collections, you can also ask for help on the objects within them, creating a kind of hypermedia help.

This Seminar now can be used by the student in several ways. The student can simply go through the activity suggested by the text in the seminar and then

BEST COPY AVAILABLE

delete it. The student might use the seminar as a tool to modify or derive some information. The student can add his own personal notes in the seminar and leave it iconified in the document for future reference. When opening the document at a later date, that seminar will help to refresh his mind to the subject matter. Similarly, the seminar might serve as an annotation to others who may look at the document.

As with all Galleries, teachers and students can extend the explanation system simply by composing a collection of helpful components and placing that collection into the Seminar Gallery. They might begin by editing the help in existing Seminars to be more specific to their situation.

DataSpace galleries provide context sensitive help which is cross-platform, user-extensible and interactive.

Activity Context for Use—DataSpace in the Classroom

Data analysis and statistics are receiving increased emphasis in the high school mathematics curriculum in two ways:

1. Following recommendations from the NCTM, topics from data analysis and statistics are being integrated into the rest of the mathematics curriculum, from geometry and algebra through to calculus.
2. Increasing numbers of one and two-semester statistics, both regular and AP, are being offered. (The first AP statistics exam will be given in 1997.)

The DataSpace project will provide three units of material—sufficient for a one-semester, regular statistics course and providing a firm foundation for an AP course. Tentative titles for the units are: Come to Your Census, Stimulating Social Simulations—Mathematical Modeling and Inference, and How We Know What We Know—Experimental Design. These units could also be taught as part of an integrated mathematics curriculum.

In addition, we are creating small instructional modules, "unettes," designed to last from one to three hours of classroom or lab time, each focused on a single statistical concept. We hope that use of these unettes, both in conjunction with the larger DataSpace units and in courses (such as science courses) not based on our materials, will help teachers and students get started using the software effectively. Unettes will tie in closely with the help system available with DataSpace.

DataSpace will be an important tool for the statistics teacher, both in the classroom and computer lab. DataSpace will take advantage of the networking infrastructure being put in place in schools, allowing collaborative data gathering, and publishing of results on the network. With a classroom computer projector DataSpace will serve as an electronic blackboard on which students and teachers can create distributions, run simulations, make and test hypotheses, and explore timely data from sources out on the Internet.

Graphing calculators are and will continue to be important data gathering and analysis tools—their portability and accessibility make them excellent classroom tools. However, the graphing calculator's small screen and modest computing power limit their usefulness in exploring the conceptual underpinnings of data analysis and statistics. DataSpace will allow easy transfer of data to and from graphing calculators so that students can begin an investigation using one tool and continue it with another.

Future Plans —The Promise of Component Software Applied to DataSpace

Over the next two and a half years, the DataSpace project will be designing, implementing, and testing software and curriculum. We will be discovering the extent to which component technology can leverage our team's limited resources toward meeting our ambitious goals.

- Components allow us to use others' technology appropriately so we do not have to reinvent it. In particular we plan to use Sketchpad's dynamic geometry and Ron Avitzur's algebraic expression editing from NuCalc[1] as part of DataSpace. OpenDoc, our component platform of choice, includes parts for text editing and access to the Internet, both very useful in DataSpace.
- Every software development project must find a design in which portions of the software are broken into relatively isolated pieces. Without this factorization, the program becomes hopelessly complicated. We are finding, in our early experiments with OpenDoc, that component technology is helping us to understand and enforce the factorization that we need.
- We will be to test individual components in isolation from the rest of the software. (Of course we also must test the ways the components work together.)
- Extending the functionality of DataSpace should be much easier with component technology than using traditional application development.
- It may prove that OpenDoc's document-centered user model is more congenial for users than the current application-centered model.
- DataSpace components should be usable in other learning contexts.
- OpenDoc is built on top of technology that promises to allow relatively painless to distributed components; i.e. components that are talking to each other over the network. We hope to leverage off distributed components to allow DataSpace users to share data and access remote collections of data.

Text and images © 1996 Key Curriculum Press

[1] Avitzur, Ron, Graphing Calculator™, ©1994.